FRANCIS JEFFRY PELLETIER and RENÉE ELIO

THE CASE FOR PSYCHOLOGISM IN DEFAULT AND INHERITANCE REASONING

ABSTRACT. Default reasoning occurs whenever the truth of the evidence available to the reasoner does not guarantee the truth of the conclusion being drawn. Despite this, one is entitled to draw the conclusion "by default" on the grounds that we have no information which would make us doubt that the inference should be drawn. It is the type of conclusion we draw in the ordinary world and ordinary situations in which we find ourselves.

Formally speaking, 'nonmonotonic reasoning' refers to argumentation in which one uses certain information to reach a conclusion, but where it is possible that adding some further information to those very same premises could make one want to retract the original conclusion. It is easily seen that the informal notion of default reasoning manifests a type of nonmonotonic reasoning. Generally speaking, default statements are said to be true about the class of objects they describe, despite the acknowledged existence of "exceptional instances" of the class. In the absence of explicit information that an object is one of the exceptions we are enjoined to apply the default statement to the object. But further information may later tell us that the object is in fact one of the exceptions. So this is one of the points where nonmonotonicity resides in default reasoning.

The informal notion has been seen as central to a number of areas of scholarly investigation, and we canvass some of them before turning our attention to its role in AI. It is because ordinary people so cleverly and effortlessly use default reasoning to solve interesting cognitive tasks that nonmonotonic formalisms were introduced into AI, and we argue that this is a form of psychologism, despite the fact that it is not usually recognized as such in AI.

We close by mentioning some of the results from our empirical investigations that we believe should be incorporated into nonmonotonic formalisms.

1. DEFAULT AND NONMONOTONIC REASONING

Default reasoning occurs whenever the evidence available to the reasoner does not guarantee the truth of the conclusion being drawn; that is, does not deductively *force* the reasoner to draw the conclusion under consideration. ('Force' in the sense of being required to do it *if* the reasoner is to be logically correct). But nonetheless the reasoner does draw the conclusion. For example, from the statements 'Linguists typically speak more than three languages' and 'Kim is a linguist', one might draw the conclusion, by default, 'Kim speaks more than three languages'. What is meant by the phrase ‘by default’ is that we are justified in making this inference because we have no information which would make us doubt that Kim was covered by the generalization concerning linguists or would make us think that Kim was an abnormal linguist in this regard. Of course, the inference is not *deductively* valid: it is *possible* that the premises could be true and the conclusion false. So, one is not *forced* to draw this conclusion in order to be logically correct. Rather, it is the type of conclusion that we draw “by default” – the type of conclusion we draw in the ordinary world and ordinary circumstances in which we find ourselves.

The example just given perhaps makes it seem obvious that default reasoning occurs in certain circumstances, namely those in which we are given *explicit* statements of typicality or normality or usualness or “for the most part” (‘Linguists typically/normally/usually/mostly speak more than three languages’). But not only do these explicit statements of typicality involve default reasoning, but so too do some ordinary statements without any restrictions. For example, such statements as ‘Birds fly’, ‘Cigarette smoking causes cancer’, ‘Ravens are black’, ‘Mary smokes a cigarette after dinner’, ‘Frenchmen eat horsemeat’, ‘Telephone books are thick’, and many others, are said to be statements involving default reasoning in that they tolerate exceptions. (A statement “tolerates exceptions” if it is true despite the existence of instances that the predicate of the statement does not truly characterize: non-flying birds, thin telephone books, instances of non-after-dinner-cigarette-smoking behavior by Mary). It has been argued (Schubert and Pelletier 1987) that there is no upper number or percentage of exceptions which these statements can allow and still be true. An alternative explanation might be that such statements are “strictly speaking” or “literally" false, but we somehow understand and use them as if they had an explicit ‘usually’ or ‘typically’, etc. We will not here consider this alternative explanation – it has not garnered a very favorable reception in the literature with which we are here concerned. It makes even the classic “Snow is white” be literally false! (For further discussion, see Krifka et al. 1995).

Formally speaking, the term ‘non-monotonic reasoning’ refers to argumentation in which one uses certain information (the *premises* of the argument) to reach a conclusion, but where it is possible that later adding some further information to those very same premises (i.e., adding another premise to the existing premises of the argument) could make one want to *retract* the original conclusion. (Sometimes this might even make us wish to conclude the opposite of the original conclusion.) Importantly, this retraction of the original conclusion is *not* accompanied by a desire to retract any of the original premises. It is this retraction-of-conclusion-without-concomitant-desire-to-retract-a-premise which sets non-monotonic reasoning apart from classical monotonic reasoning. Put symbolically, it is a case of non-monotonic reasoning if one is willing to make the inference $\{P_1, P_2, \ldots P_n\} \therefore C$ but is unwilling to make the inference $\{P_1, P_2, \ldots P_n, P_{n+1}\} \therefore C$. The catch-phrase that is the hallmark of non-monotonic reasoning is "that new information makes one withdraw previously-made inferences".[1]

It is easily seen that the informal notion of default reasoning manifests a type of non-monotonic reasoning. In the above example, for instance, we concluded that Kim spoke at least three languages. But were we to add to our list of premises the further fact that Kim graduated from NewWave University, which we know has revoked all language requirements, we then would wish to withdraw the earlier conclusion. Thus, default reasoning is a species of non-monotonic reasoning. More generally speaking, default statements are said to be true about the class of objects they describe, despite the acknowledged possible existence of "exceptional instances" of the class. In the absence of explicit information that any particular object is one of the "exceptional instances", we are enjoined to apply the default statement to the object. However, further information may arrive telling us that this object in fact is one of the "exceptional" ones. This is where non-monotonicity resides in default reasoning.

Various philosophers and logicians have tried to give an account of how ordinary people perform default reasoning. In general, the view taken by most of these earlier philosophers was that the people in question "jumped to conclusions" which were not really logically justifiable, but which were required to be made on the basis of insufficient information.[2] A natural outgrowth of this attitude toward people's use of non-monotonic reasoning is that such reasoning "really" is bad deductive reasoning – perhaps justifiable on the grounds of having to "get on with it" in the face of limited resources such as time and energy, but bad nonetheless. This attitude is quite common in the philosophical literature on defeasible reasoning. A similar attitude toward probabilistic reasoning is also quite common in some of the psychological literature. For example, Tversky and Kahneman (1983) found that people will assign a lower probability to the proposition *There will be a very severe earthquake in costal California in the next three years* than they do to the proposition *There will be a very severe earthquake in costal California in the next three years and property damage will exceed $500 million because numerous houses will fall into the ocean*. Yet of course the second proposition is a conjunction with the first proposition as one of its conjuncts, and therefore the second proposition cannot have a higher probability than the first. Tversky and Kahneman view people as *ignoring* relevant information and *making mistakes* in their probabilistic reasoning.

But non-monotonic reasoning, in our view, is *not* an example of bad deductive reasoning. It is not an example of doing something wrong. Following many theoreticians in artificial intelligence (AI), we believe that it is *correct* to make such inferences. It is *not* a mistake on people's part, nor is it a matter of "having to do *something*, anything, in the face of insufficient information". Rather, it is right and proper to make such inferences: not only is this what people *in fact* do, but it is what people (and artificial agents) *ought* to do.

2. USES OF DEFAULT REASONING

As remarked above, many logicians question whether the notion of non-monotonic logic is coherent. But interesting as this question is, we do not propose to investigate the philosophical question of whether non-monotonic logic makes any "deep" sense. Instead we lay out a number of academic realms that have employed default reasoning, with an eye to demonstrating just how widespread the phenomenon really is. We think that the independent appearance of this mode of reasoning in widely-divergent fields, with little evidence of any cross-fertilization, shows how pervasive it is and how deeply this sort of reasoning seems to be embedded in our explanations of nature and of human nature. We find it rather surprising that there has been so little cross-fertilization amongst these different areas, and would urge researchers to seek out the investigations carried out in fields other than their own. Here we only mention, rather than characterize fully, certain areas in which the notion of default reasoning has been investigated.

2.1. *Ethics*

"Everyone knows that it is wrong to tackle a stranger on the street. Except, of course, if he is running away from a crime. Unless a policeman is already chasing him and is about to capture him. But if you see that the criminal is really going to escape, then you should tackle him. Unless you know that it is a corrupt policeman ...". Ethical reasoning almost always has this sort of default character: there are non-standard circumstances which can be added to premises already accepted and which will make us withdraw our earlier conclusions concerning what we should do. Another related ethical example is: "You should not steal. But if you do, then you should make restitution". Formal ethicists have long been interested in developing "deontic logics" – logics of obligation and permission. If they honor the types of intuitions just mentioned, then they will be nonmonotonic logics, exemplifying default reasoning.

In one of the popular terminologies in ethics (due to Ross 1930), ethical "laws" only give *prima facie* obligations; they are always accompanied by *ceteris paribus* ("other things being equal") conditions. Similarly we only have *prima facie* rights. These obligations, duties, and rights are real, and they are warranted by our ethical code. Yet nonetheless they are defeasible: further factual information might come in to demonstrate that we do not have the obligation, duty, or right. If we could specify *all* the *ceteris paribus* clauses – that is, make the default rule become absolute – then we would be able to derive an 'ought' statement from an 'is' statement; i.e., we would be able to derive an ethical statement from what in fact is happening. And this is something that few scholars think is possible. (See Searle 1964 for the argumentation relevant to this derivation.) Ethics is probably the area in which default reasoning has been investigated for the longest time.

2.2. *Generics and Habituals*

In linguistic semantics, researchers have long been interested in the conditions under which such sentences as the following will be true: 'Bears with blue eyes are intelligent', 'Dogs chase cats', 'Mary jogs to work', etc. The first two of these examples are called generics, since they talk about genera (Bears, Dogs, Cats) and do so in a general or generic way. The third sentence is called a habitual, since it reports a habit of Mary or a habitually occurring episode. Obviously generics and habituals are related, and it is especially when considering the truth-conditions of such sentences that one can see the relationship. In fact, *not* all dogs chase cats; and sometimes Mary does *not* jog to work (sometimes she isn't going to work and hence isn't jogging to work; other times she is going to work but decides to walk or ride). Because such objects and facts are "exceptional" in the terminology developed above, they do not by themselves undermine the truth of the generic or habitual sentences. Thus they can give rise to the hallmark of default reasoning: we *assume* that an object is *not* "exceptional", and hence use the generic and habitual sentences to draw conclusions about the object. But we are prepared to withdraw these conclusions upon finding that the object *is* "exceptional". Understanding the semantics of such sentences is a major research effort in linguistics, and it is clearly quite closely related to other work done in different fields concerning default reasoning. (For a survey of work done in the field, see Carlson and Pelletier 1995.)

2.3. *Philosophy of Science*

In the traditional conception of science, laws of nature are viewed as universally true, exceptionless statements of nomic or lawlike necessity. But of course it has long been pointed out that scientific laws are always accompanied by a *ceteris paribus* clause. And no scientist believes that one can give an exhaustive enumeration of all the "other things" which must "be equal" in order for the law to be universally applicable. This in turn has led to the view that the laws of science *aren't* literally true; and this in turn leads to a philosophical position known as "anti-realism" – that our scientific statements about nature do not really describe an independent reality. A "realist" counter to this argument is that laws of nature, even with their *ceteris paribus* clauses, *do* describe reality. They merely do it in a default manner, but they do it nonetheless. Clearly, the realist position described here has at bottom the same sort of puzzles that all the previous areas have delineated: how can a natural law be true and still allow for exceptions? How can we draw conclusions concerning particular objects from general laws and later retract them if we find that the object is "exceptional"? The challenge in philosophy of science is to give philosophical and logical sense to such a conception without thereby admitting that we are not "talking about reality". For one side of this debate, the antirealist side, see Cartwright (1983, 1989).

2.4. *Conditionals and Counterfactuals*

A conditional statement is (prototypically) one of the form 'If X, then Y'. A counterfactual conditional is one in which the antecedent, X, is presumed or presupposed to be false. It has been a puzzle for quite a while in philosophical logic as to what the conditions are under which such statements are true. For instance, Lewis (1973) pointed out that 'If kangaroos had no tails, they would fall over' was true while 'If kangaroos had no tails but used crutches, they would fall over' was false, but that 'If kangaroos had no tails but used crutches that were sawed in half, they would fall over' was true. And it seems that such an alternation of true and false could be continued indefinitely by suitably choosing antecedents for the conditionals. However, this is really quite puzzling: the antecedent of the first statement in this list was 'kangaroos have no tails'. And as a special case of this, that is, as one of the ways this might happen, we have 'kangaroos have no tails but they use crutches'. Yet this special way of kangaroos having no tails does *not* support the consequent of the conditional ('they fall over'). It seems clear that kangaroos using crutches is an "exceptional" way of kangaroos having no tails ... and it is for that reason that we do not draw the same consequent. Similarly, having sawn crutches is an "exceptional" way for kangaroos to have no tails and use crutches. This leads formally to nonmonotonic logic because from a premise 'Kangaroos have no tails' we would conclude 'Kangaroos fall over', from the initial premise mentioned above. Yet if we were also given the premise 'Kangaroos use crutches' then we no longer draw that conclusion even though we do not retract any previous premises.

2.5. *Relevant Logic*

Relevant logic grew out of a dissatisfaction with the manner in which classical logic treated the notion of being relevant. There are a number of such unintuitive results in classical logic. For instance, if we are given 'John will go to Sumatra' as premise, we should not thereby be justified in concluding 'if Martin Luther King was assassinated then John will go to Sumatra'; for, our intuitions tell us that the information about King is simply not "relevant" to John. Yet classical logic counts this inference as valid, as it does the inference that from a contradiction *everything* follows, no matter how unrelated it is to the contradiction. And there are many other inferences in classical logic which are equally objectionable, according to relevant logicians. The movement in relevant logic has been going on for some 40 years now and has developed many sub-strands within it. A salient feature of most of these developments is that many of the inferences they find objectionable are closely related to those which give rise to interest in default reasoning. It therefore seems that a close inspection of the notions of "exceptional" in default reasoning and "relevant" in relevant logic will allow them to shed light on each other. (Relevant logic bibles are Anderson and Belnap 1975; and Anderson et al. 1992.)

2.6. *Prototypes, Stereotypes, and Schemata*

There is a large empirical literature in cognitive psychology on notion of prototypes: their genesis, the nature of their internal representation, and their use in making judgments about novel occurrences in the world. There is no standard agreement as to what a prototype is (and in some cases, arguments that much of what this notion is used for could be equally well accounted for by analogy to specific example cases). The debates on these matters can become very involved. Still, both empirical data as well as general intuition support the idea that "prototype" might be a useful cognitive construct. A prototype may be a best example or best set of examples of some category; it may never have been sensed, nor even really exist in the world. It might instead be a constructed representation of information about the examples that have been encountered. Prototypes are distinguished from stereotypes (see Putnam 1970, 1975; Johnson-Laird 1983, 187–197), which are sometimes defined as a set of properties "typically associated" with a mental category.

Although defining or discerning the nature of the mental representation of prototypes or stereotypes is a complex theoretical and empirical matter (Medin and Smith 1984), we do not need to commit to one resolution or the other of this matter to draw the connection to non-monotonic reasoning. By definition a prototype or stereotype is *not* universally true of, or adequate to, all the members of the category. According to this theory, when one reasons about whether Simba, the lion, is ferocious, one allegedly consults one's prototype or stereotype of Lion to see whether it is or is not ferocious. And unless we know something special about Simba which would make it "exceptional", we conclude that it follows the prototype or stereotype. As can be seen, the use of prototypes in this manner – leaving open the possibility that any inference based on the prototype may have to be retracted in the future – is non-monotonic. Indeed, it is very close to the application of a default set of expectations that are in some way or other known not to hold for an entire class. This is particularly true when prototypes are viewed from an inductive perspective: it is the use of some induced knowledge that is true, at best, only about the members encountered to date. (Classic works are Rosch 1978; Smith and Osherson 1984.)

2.7. *Causal Reasoning*

Reasoning about causation is pervasive in ordinary discourse. One way the topic arises is if one attempts to predict what will happen at a given time when one has a theory which causally explains the relevant phenomena. Because causal laws are often equipped with escape clauses (*ceteris paribus* clauses) and because a phenomenon might fall under two competing causal laws, we might conclude that a certain phenomenon will happen and yet we later discover that the escape clause is active or that another law overrides the law to which we are appealing and we retract our conclusion. Thus causal reasoning of this sort is defeasible.

Another source of defeasibility in causal reasoning is in the attempt to infer causation from observations. Having seen a number of instances in which X followed after Y, and none where X followed when Y was absent, we sometimes wish to infer that Y causes X. Yet it is clear that such an inference is defeasible: further evidence could make us wish to retract this causal conclusion.

The general issue of constructing a formal calculus in which one can reason from causal statements to particular occurrences and in which one can reason from observational statements to statements of causation has been extensively addressed in philosophy, some of the classical statements of formal theories being Good (1961/1962), Suppes (1970), and Fetzer and Nute (1979). It is also a research area in AI to find a formalism in which statements about causation can be couched and predictions about what will and what will not change as the result of actions by people or robots. This area has garnered an immense amount of literature, in which some of the early, foundational works are McCarthy and Hayes (1969), McDermott (1982), McCarthy (1986), Hanks and McDermott (1986, 1987), Shoham (1990), Goebel and Goodwin (1987), Jachowicz and Goebel (1997), and Pearl (1988, 2000).

2.8. *Diagnosis*

'Diagnosis' can be viewed as finding a difference between what is expected and what is actually occurring in some system. The system might be an electronic circuit, or it may be a human body. Formal accounts of diagnosis are often divided into two sorts: one where we have a description of the system together with observations of the system's behavior and the other where we have "heuristic" information of the sort 'When the system exhibits this behavior, then in 80% of the cases the following component has failed'. The first sort of diagnosis is called 'Diagnosis from First Principles', and is the sort to which we draw attention (central works include Davis 1984; de Kleer and Williams 1987; Reiter 1987). In this framework, if the observation conflicts with the way the system is meant to behave, the goal is to determine those system components which, when assumed to be functioning abnormally, will explain the discrepancy between the observed and correct system behavior. Reiter's approach to diagnostic reasoning in this system is a form of default reasoning: it can happen that none of the diagnoses we infer about a system survives a new observation of that system. Thus, a conclusion based on certain observations might be that such-and-such fault or abnormality is present in the system, yet a further observation could lead to retraction of this conclusion without withdrawing belief in any of the previous observations. (Reiter 1987 gives a formal characterization of the relationship between nonmonotonic logic and his diagnostic analysis. In Poole et al. (1987) a default logic theorem prover is used to compute diagnoses.)

2.9. *Reasoning in Social Sciences*

One of the differences between the social and natural sciences has always been seen as the former's having laws which are not "strict" or "universal" in the same way as the latter's.[3] Some attribute this to a presumed statistical nature of social science laws (see Salmon 1990 for discussion) and others attribute it to social sciences not "really" being sciences. However, another tack that might be taken is that social-scientific laws are default laws. In accordance with this, (Janssen and Tan 1991, 1992) point out that some of the apparently "nonscientific" nature of economic laws proposed by Milton Friedman, (e.g., the Permanent Income Hypothesis, see especially Friedman 1957), can be accounted for in this manner – perhaps deflecting some of the charges that have been leveled against Friedman's economic theory. (For discussion on both sides, see Tobin 1971; Hirsch and de Marchi 1986; Hammond 1988; Hausman 1989). Janssen and Tan reformulate Friedman's theory with three variables and a variety of equations relating them, making it become analogous to the type of system discussed above under "diagnosis". The observations which contradict the statements of the proper operation of the laws (those studies which show that in certain specific instances the predictions of the Permanent Income Hypothesis do not come to pass) are treated as "faults" in the theory. Janssen and Tan reformulate the Permanent Income Hypothesis with "abnormality predicates" within Reiter's (1980) default logic and show that using the diagnostic techniques of Reiter (1987) we can restore consistency to the overall theory. It seems that a similar tactic could be applied in many places in the social sciences.

2.10. *Judgment under Uncertainty*

Probabilistic reasoning is a clear example of nonmonotonic reasoning, at least in those cases where one allows that a conclusion should be drawn whenever its probability is greater than r, for some r less than 1 (certainty). For in such a case we would be able to draw a conclusion on the basis of our evidence (because the conclusion has a probability greater than r, on the basis of the evidence), but then it can happen that we receive further information which now makes this old conclusion have a probability less than r, on the basis of what is currently all the evidence. Given the premises that the probability of a Scandinavian being a Muslim is less than 0.02 and the certainty that Olaf is Scandinavian, we conclude that Olaf is not a Muslim (the probability of this is greater than our cutoff, r). Yet if we then discover (with certainty) that Olaf is traveling to Mecca and we know that the probability of someone going to Mecca and not being Muslim is less than 0.03, we would withdraw our previous conclusion about Olaf not being Muslim.

The issue of how people actually reason with their probability judgments (and other related methods) has been studied in both Psychology and in Management Science. In the former sort of study, much attention has been paid to how people allegedly *misuse* the probability calculus (see

Tversky and Kahneman 1983; Nisbett et al. 1983). In the latter sort of study the goal has been to study (a) how decisions are made by considering what information is employed by the decision-makers, (b) how good are people's decisions and the cognitions that underlie them, (c) how to improve decisions when they are deficient.[4] The general thrust of such studies is to investigate how new information impacts on established beliefs. Thus there will be at least a portion of this research which overlaps default reasoning: at a certain time, a person's standing beliefs leads to the conclusion C, yet upon receiving new information the person updates the beliefs so as to no longer include C. Much of the research is to determine the conditions under which people update beliefs in one way vs. some other way (what sort of topics, what sort of educational background, etc.). This sort of belief revision is clearly a close relative to default reasoning.

2.11. *Implicatures*

'Implicature' is a term coined by Grice (see especially his 1975, 1978, 1981) to designate an item of information that a speaker wishes to convey but which is not part of the "literal meaning" of the utterance, that is, not part of what is literally said, in a strict sense of the phrase. For example, if a colleague is seeking a particular issue of a journal and you tell her that you own a copy, you have implicated that it is available for her to see. Now, you haven't *said* that, literally; rather you have *implicated* it. Indeed, you could continue the conversation by mentioning that although you own a copy of it that it has been borrowed by some other colleague who took it on sabbatical to Tasmania, thereby "canceling" the implicature that your journal is available for her. And doing so would not in any sense be a contradiction of what you had said – you would not be going back or giving up any of your previous statements. It can be seen that implicatures amount to a kind of default reasoning: when a speaker makes an utterance which has an implicature, the hearer is entitled by default to infer that implicature. Still, the speaker could go on to cancel the implicature by giving more information, and the hearer would withdraw the inference without withdrawing any of "what was said". (The general default nature of implicatures has been followed up by Clark and Haviland 1977; Levinson 1983; and Sperber and Wilson 1986 – among many others.)

2.12. *Linguistics*

In recent years default mechanisms have been increasingly employed in different aspects of formal linguistics. The idea of representing information about inflectional morphology as a matter of there being "default cases" vs. "exceptional classes" (and "exceptional subclasses") is pursued in Evans and Gazdar (1996). And more generally about the lexicon there is work at building defaults into the unification of feature structures (Lascarides et al. 1996). Additionally, there is work in trying to explain certain discourse phenomena such as anaphoric underspecification (of both pronouns and temporal reference) by means of defaults (Lascarides and Asher 1993). Thomason (1997) contains a survey of linguistics-related work.

2.13. *Natural Logic*

Natural logic and its cousin natural language metaphysics, as sciences, are the investigation of formal principles of inference and their ontological presuppositions as they actually occur in natural language. One of the basic premises of natural logic is that the process of regimentation – that is, the process of representing natural language statements in some artificial language – can introduce unintended features which can then interact in such a way as to lead to conclusions which we would intuitively judge as wrong. For example, the regimentation of 'Pegasus is a winged horse ridden by Belaraphon' into first-order logic commits us to the existence of Pegasus. There are many such hidden commitments in the various languages of regimentation in use, and, to the extent that such commitments are unrecognized by the its users, they will be wrong about "natural reasoning". One of the areas of natural logic that has been seriously studied concerns "hedges": expressions which modify or soften or exaggerate how some other statement should be taken. For instance, the phrase 'technically speaking' in the sentence 'He is technically speaking the departmental chairman' is a hedge, and it has the effect of modifying the standard understanding of what it is to be a departmental chairman (for instance, he has been appointed but never does any of the work, or he is a mere figurehead, etc.) As can be seen, this hedge appeals to some sort of stereotypical information about chairmen and says that the person does not satisfy it. But it does *not* point to any *specific* piece of stereotypical information. This means that any inference one might draw from the statement is defeasible: we are not surprised when it turns out that we picked on the wrong inference. For instance, we might conclude from that sentence that the person does none of the departmental work. But if someone explained to us that this wasn't what he meant, that instead he meant that there is some other power behind the chairmanship, we would retract our conclusion. (See Lakoff 1972, 1973 for this viewpoint. Also compare Braine 1978; Henle 1962; and Macnamara 1986.)

2.14. *Cognitive Science*

A certain argument in cognitive science goes like this (Oaksford and Chater 1991; Garnham 1993; Chater and Oaksford 1993): Defeasible inference is a pervasive feature of human cognition, something that must be given a central place in any account of the science of the mind. Yet the only "formal" or "mechanical" or "proof theoretic" or "algorithmic" account of defeasible reasoning is that given by computer science in terms of nonmonotonic logic, and this account does not have a computationally-tractable proof theory. The centrality of default reasoning tells us that the failure of a computational account hits at the very center of what the science of the mind must be. The "dominant paradigm" according to which cognition is explained by recourse to a mechanized proof theory operating on a Language of Thought (Fodor 1975, 1983; Pylyshyn 1984; Fodor and Pylyshyn 1988) falters on the computational intractability of default reasoning. Thus the computational theory of mind, so-called, must be wrong.

2.15. *Knowledge Representation*

A trend in artificial intelligence over the last few decades has been to investigate the possibility of constructing "knowledge bases" – which are envisaged as a type of database, but where there is much commonsense reasoning ability built in (Reiter 1992; Davis 1990; Levesque and Lakemeyer 2001). This trend pursues an analogy with actual human information storage, where not all of our information about the world is stored in an explicit form in our minds, but we can somehow generate it as needed. For instance, few of us have stored the fact that there are support beams under every university building, yet we each know this nonetheless. A goal of those who would construct knowledge bases is to mimic this manner of storage. It is clear that a knowledge base must have the ability to draw default conclusions, that is, to make inferences to a conclusion using a certain set of information but to be prepared to retract that conclusion when more information comes to the fore (without withdrawing any of the old information that was initially used). It is in fact this area of AI that has seen the most sustained and serious attempts to develop formalisms to account for default reasoning. And it is to this area that we should look when we wish to find serious portrayals of the factors involved with this type of reasoning. And it is to this arena that we will direct our attention later in this paper.

3. PSYCHOLOGISM IN REASONING

Although default reasoning is a type of reasoning and therefore it shares certain properties with other types of reasoning such as deductive reasoning, there is at least one important difference: deductive reasoning has a "normative standard" that is "external" to people whereas default reasoning has no such external normative standard ... or so we will argue.

In deductive logic, the external normative standard for a good argument is that of truth-preservingness: the truth of the premises will guarantee the truth of the conclusion. And we can see that it is possible for people to fall short in determining this: they might, of their own accord, draw conclusions that are not thus guaranteed and fail to draw ones that are guaranteed; they might fail to recognize correct and incorrect arguments for what they are; and they might even deny that an argument is valid (or invalid) when it is recommended to them as valid (or invalid). All this is possible because there exists the independent-of-people, external standard of correctness against which we can make these evaluations. And similar remarks could be made about mathematics: there is an independent, external standard of correctness, so it makes sense to claim that some people commit errors in their mathematical reasoning. Frege's influential review of Husserl (Frege 1894, see also Frege 1884) persuaded almost all theorists that psychologism was false of mathematics and logic (in particular), so that logic was seen not to be a "subjective" enterprise but instead it concerns objective relations amongst propositions, predicates, and terms. Nowadays, one would be hard-pressed to find anyone who holds psychologism with regards to logic, mathematics, geometry, and the like.[5] As a consequence, when we investigate how people actually reason in these realms, our conclusions must be different than if we believed in psychologism for that field; for according to psychologism, people (as a whole) *cannot* make mistakes about the field. If (almost) everyone reasons in such-and-so way concerning logic, then by definition (according to psychologism) such-and-so *is* logic. It is only if we reject psychologism in logic that we can say that subjects *make mistakes* in these areas. We discover that subjects *make more mistakes* in reasoning with Modus Tollens than with Modus Ponens, for example (see Evans 1987 for the relevant data); and this is a possible discovery only when psychologism is rejected.

But the case of default reasoning is different. Here there is no external standard of correctness other than what people actually infer. Of course, an external standard *could* be invented – for example, an AI medical diagnosis system might work by using "default principles" to come up with diagnoses of diseases based on reported symptoms and medical history.

And it would make “correct default inferences” only to the extent that it correctly diagnosed the diseases.[6] But this is not the type of situation we are faced with in the default reasoning cases of interest. We think the interesting cases are more akin to determining how North Americans distinguish green from blue, for example, where there is no real concept of right and wrong in the task. Some people draw the boundary in one place, some in another place; we can tell whether there are identifiable subgroups who all draw it at one location while other subgroups draw it in another place; and we can tell whether someone is different from the majority in where s/he draws the line. But there is no real notion of “draws the blue/green boundary incorrectly”; all that exists is how people in fact do it – there simply *is no other standard*. This attitude is called “psychologistic” because it locates the object of the study (or the normativity of the theory) in the psychology of people and denies that there is any “external standard of correctness” for the field. And it is this attitude that we wish to endorse for default reasoning.

Our claim is that default reasoning is psychologistic, that is, what is and isn’t correct default reasoning is *defined by* what people do. To be clear about this, we emphasize that our idea allows or even requires there to be a notion of “mistaken default inference”, but such a notion will itself need to be defined in terms of people’s general performance. For example, we might say that a person is making a mistaken default inference if it is at odds with the generally-accepted consensus about the inference – rather like we could say that Smith does not draw the blue-green boundary in the same way as his neighbors and therefore this could lead to trouble in conversation (and house-painting). We might also say that a default inference is mistaken if it contradicts what that person acknowledges on other grounds as the correct conclusions to draw. (Or at least, the person would have to withdraw one or the other of the methods s/he used to generate the conclusions s/he admits to being in conflict.) We therefore do not find ourselves open to the charge that “anything goes” or that “there are no standards” and “no one’s inferences can be corrected”, if psychologism is correct about default reasoning.

As we remarked above, we think the most serious attempts to characterize the formal properties of default reasoning are to be found in the knowledge representation literature. We wish to draw out the psychologistic presupposition that lurks behind their work, and employ that as a justification for our empirical investigations. In Pelletier and Elio (1997) we canvassed numerous works by researchers in the field in order to discover their underlying justification for their investigations into nonmonotonic logic. Here four sample quotations (out of very many) from this study:

> A key property of intelligence – whether exhibited by man or by machine – is *flexibility*. This flexibility is intimately connected with the defeasible nature of commonsense inference ... we are all capable of drawing conclusions, acting on them, and then retracting them if necessary in the face of new evidence. If our computer programs are to act intelligently, they will need to be similarly flexible. A large portion of the work in artificial intelligence on reasoning or deduction involves the development of formal systems that describe this process .... Unfortunately, the mathematical work on inference has only recently become concerned with flexible inference of the sort we are discussing. Conventional deductive inference has a property known as *monotonicity* .... In addition to applications to the understanding of common-sense reasoning, nonmonotonic reasoning also has been shown to be important in other areas. There are applications to logic programming, to planning and reasoning about action, and to automated diagnosis. (Ginsberg 1987)

> It has been generally acknowledged in recent years that one important feature of ordinary common-sense reasoning that standard logics fail to capture is its *nonmonotonicity*. ... Autoepistemic logic is intended to model the beliefs of an agent reflecting upon his own beliefs. ... We are trying to model the beliefs of a rational agent .... An autoepistemic logic that meets these conditions can be viewed as a competence model of reflection upon one's own beliefs. ... It is a model upon which the behavior of rational agents ought to converge as their time and memory resources increase. (Moore 1985)

> It is commonly acknowledged that an agent need not, indeed cannot, have absolute justification for all of his beliefs. An agent often assumes, for example, that a certain member of a particular kind has a certain property simply because it is typically true that entities of that kind have that property. ... Such default reasoning allows an agent to come to a decision and act in the face of incomplete information. It provides a way of cutting off the possibly endless amount of reasoning and observation that an agent might perform .... (Selman and Kautz 1989)

> Most of what we know about the world, when formalized, will yield an incomplete theory precisely because we cannot know everything – there are gaps in our knowledge. The effect of a default rule is to implicitly fill in some of those gaps by a form of plausible reasoning .... Default reasoning may well be the rule, rather than the exception, in reasoning about the world since normally we must act in the presence of incomplete knowledge .... Moreover, ... most of what we know about the world has associated exceptions and caveats. (Reiter 1978)

We see in all these quotations the view that "traditional" or "mathematical" or "formal" logic is inadequate for the task of describing ordinary human activities and commonsense reasoning; rather, these ordinary human activities rely upon people's ability to employ some form of non-monotonic, default reasoning. In consequence, progress on the development of intelligent artifacts (robots and the like) will not advance until more research is done into non-monotonic reasoning and the ways it interacts with other intelligent activities. It is this feature of "non-monotonic flexibility" that needs to be instantiated in these artifacts, in order for them to be correctly viewed as intelligent.

This is clearly a psychologistic attitude – what is a good or valid ability in default inference is defined in terms of what allows people to "get along in the world". The point we wish to emphasize and to which we wish to draw the reader's attention is this: *Despite the acknowledgement by the artificial intelligence community that the goal of developing non-monotonic systems owes its justification to the success that ordinary people have in dealing with default reasoning, there has been no investigation into what sorts of default reasoning ordinary people in fact employ*. Instead, artificial intelligence researchers rely on their introspective abilities to determine whether or not their system ought to embody such-and-so inference. And even the so-called "Benchmark Problems" that we will discuss in the next section were formulated with absolutely no regard to whether ordinary people in fact do reason in the way prescribed! Given the central place that these Benchmark Problems occupy in the field – they are the minimal abilities that any artificial system must embody – a crucial research question is whether or not non-monotonic reasoning as conceived by the non-monotonic reasoning community actually conforms to the promises and goals initially held out for it, especially those promises that it would yield up the sort of reasoning that people actually engage in. Principal to this is the question of the extent to which the non-monotonic community has accurately characterized "ordinary", "commonsensical" reasoning. After all, the example non-monotonic arguments cited in the literature were all invented *ex nihilo* by theorists. None of them empirically investigated the extent to which real "ordinary, commonsensical reasoning" agreed with the examples. Yet, it was precisely such an agreement that was the entire *raison d'etre* for the enterprise. We therefore pose the question: Do people actually reason in the manner prescribed by the non-monotonic logic community? That is, do they agree with the conclusions judged valid by the non-monotonic logic community?

4. THE NON-MONOTONIC BENCHMARK PROBLEMS

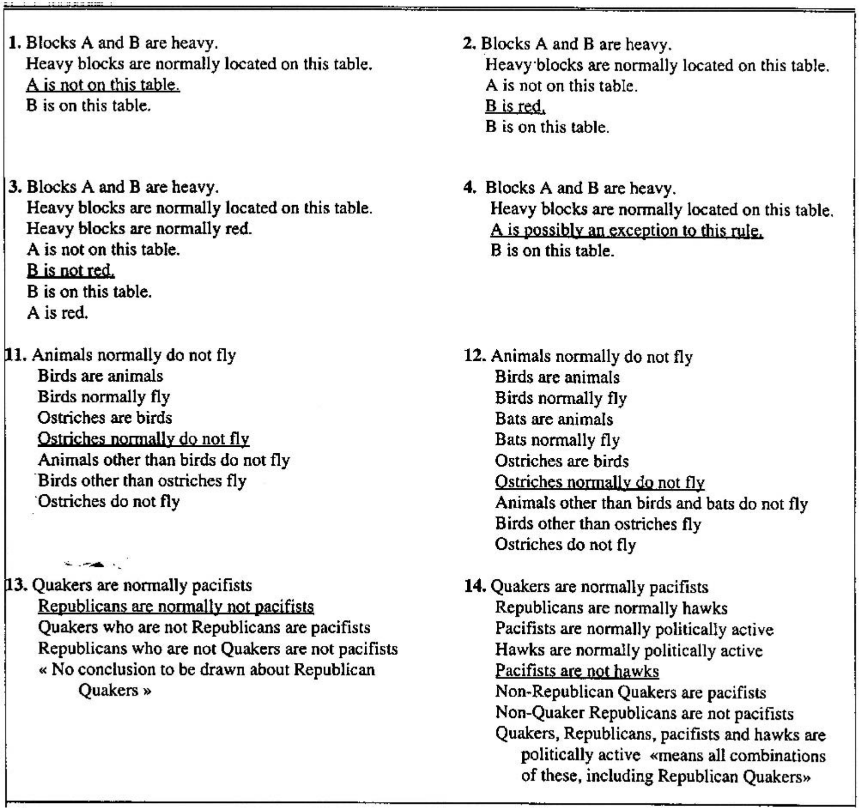
Despite the fact that all the researchers in the field appeal to the goodness of human performance as their justification for employing default reasoning mechanisms in artificial agents, the truth is that none of them in fact have ever investigated how people actually employ default reasoning. A consequence of this is that the different proposed formalisms validate different sets of inferences, and there is no agreed-upon method to decide which inferences should be sanctioned as "really legitimate". Recognizing that there was no accepted background for finding the extent of legitimate default inferences, Lifschitz (1989) set out a number of inferences that were supposed to be valid in any proposed default reasoning system. Different groups of these problems were addressed to different aspects of the default reasoning process, so that perhaps not every reasoning system needed to accommodate all problems; but for any area for which a system proposed a mechanism, it was to be able to deal at least with the inferences relevant to that area. We call these problems "the Benchmark Problems", and the accepted answers that were proposed for these problems in Lifschitz (1989) the "AI answers" to the Benchmark Problems.

Although one might accept the legitimacy of establishing a set of benchmark problems in this way because these are the areas in which default reasoning is to be employed by artificial agents, one might nonetheless wonder about the legitimacy of determining the AI answers as a matter of agreement among the various researchers. After all, if it is true, as we argued the field of knowledge representation to be committed, that the realm of default reasoning is psychologistic and that therefore the correct answers are determined by the way "ordinary people" will (on the whole) use the method, then the fact that some elite subgroup of people think the answers should be such-and-so is not a good justification. For one thing, their opinions might be colored by how their own systems perform. More importantly, an individual's intuitions are not a reliable guide to how the population as a whole treats the phenomenon. In Pelletier and Elio (1997) we have investigated the various reasons researchers give for allowing their own intuitions to be their sole guide in this regard and for not engaging in large-scale investigations of how it is manifested in the population as a whole. We did not find any of these reasons very compelling, and recommended that researchers undertake such studies in order to determine the correct direction for their formal theories to follow.

The two areas of Lifschitz's list of Benchmark Problems that we have investigated are "Basic Default Inference" (his Problems #1–#4) and "Inheritance Reasoning" (his Problems #11–#14). Figure 1 gives them as they are stated in his article:

5. SOME EMPIRICAL RESULTS ABOUT DEFAULT REASONING

In the present paper we quickly mention some of the results of our experiments. Despite the tentative nature of the results, we think they should give default reasoning researchers pause in their confidence that they have in fact fathomed the true nature of the reasoning process they are trying to model. More details about the experimental framework and a more



1. Blocks A and B are heavy.
Heavy blocks are normally located on this table.
A is not on this table.
B is on this table.

2. Blocks A and B are heavy.
Heavy blocks are normally located on this table.
A is not on this table.
B is red.
B is on this table.

3. Blocks A and B are heavy.
Heavy blocks are normally located on this table.
Heavy blocks are normally red.
A is not on this table.
B is not red.
B is on this table.
A is red.

4. Blocks A and B are heavy.
Heavy blocks are normally located on this table.
A is possibly an exception to this rule.
B is on this table.

11. Animals normally do not fly
Birds are animals
Birds normally fly
Ostriches are birds
Ostriches normally do not fly
Animals other than birds do not fly
Birds other than ostriches fly
Ostriches do not fly

12. Animals normally do not fly
Birds are animals
Birds normally fly
Bats are animals
Bats normally fly
Ostriches are birds
Ostriches normally do not fly
Animals other than birds and bats do not fly
Birds other than ostriches fly
Ostriches do not fly

13. Quakers are normally pacifists
Republicans are normally not pacifists
Quakers who are not Republicans are pacifists
Republicans who are not Quakers are not pacifists
« No conclusion to be drawn about Republican Quakers »

14. Quakers are normally pacifists
Republicans are normally hawks
Pacifists are normally politically active
Hawks are normally politically active
Pacifists are not hawks
Non-Republican Quakers are pacifists
Non-Quaker Republicans are not pacifists
Quakers, Republicans, pacifists and hawks are politically active «means all combinations of these, including Republican Quakers»

Figure 1. Eight Problems from Lifschitz's Benchmark Problems.

quantitative presentation of the results can be found in Elio and Pelletier (1993, 1996); Pelletier and Elio (2003). In general our results on these eight problems coincided with the AI answers, although we also uncovered some interesting differences. In this paper we merely present the overall results that are at odds with the AI answers, without any detailed discussion. Our overall view is that, because of the psychologistic nature of default reasoning, these findings need to be assimilated into the nonmonotonic formalisms being recommended by AI researchers. And it has seemed to us that the only formal theory that appears to be able to do this sort of assimilation is that of Pollock (1987, 1991).

Each of Problems #1–4 concerns two objects governed by one or more default rules. Additional information is given to indicate that one of the objects (at least) does not follow one of the default rules. We refer to this as the *exception object* (for that default rule). The problem then asks for a conclusion about the remaining object, which we refer to as the *object-in-question*. For all these problems, the AI conclusion is that the object-in-question (Block B) obeys the default rule concerning location. According to the collective wisdom of researchers into nonmonotonic theories – the existence of an exception object for a default rule, or additional information about that exception object, should have no bearing on a conclusion drawn about any other object when using that rule. Extra information about the object in question itself (e.g., Block B's color) should also have no bearing on whether a default rule about location applies. And being an exception object for some *other* default rule should have no bearing on whether it does or does not follow the present default rule.

An implicit but important assumption here is that a logic for manipulating assertions of this *form* can be developed for non-monontonic reasoning *without regard for the semantic content of the lexical items*. Given these formal problems about blocks and tables, it seems easy to accept the idea that object color should have no "logical" bearing on whether a default about object location is applied. Yet it is equally easy to imagine real-world scenarios in which it makes (common) sense that an object's color might be predictive of or at least related to an object's default location (e.g., how an artist or designer might organize work items in a studio). We do not wish to confound people's *logical* abilities with their ability to "look up" information they have stored in memory. And so we would want to test them on scenarios for which they have some commonsensical "feel" but for which they have no stored information. Similar remarks hold about the Inheritance Reasoning problems: Subject's knowledge about real-world bats and ostriches, as well as Nixon, should not be allowed to influence their *logical performance*.

With regard to the Basic Default Reasoning problems, our results are as follows. First, subjects see Benchmark Problem #3 as quite different from #1, 2, and 4. This suggests that the more default rules an object is known to violate, the more "generally exceptionable" the object is, and the more likely the object is to violate other defaults. We enshrine this observation as:

Second-Order Default Reasoning: If the available information is that the object-in-question violates other default rules, then infer that it will violate the present rule also.

Others might prefer to call this the "Guilt by Past Association" rule, or maybe the "Bad Egg" principle ("once a bad egg, always a bad egg"). Secondly, we also varied the amount of "extraneous material" that claimed the exception-object and the object-in-question were similar. Our view is that subjects interpret any "extra" information as possibly giving a reason why the exception object did not obey the default rule, and perhaps an indication that this reason applies to the object-in-question also.[7] This view was supported by the data, and we enshrine it as:

Explanation-based Exceptions: When the given information provides both a relevant explanation of why the exception-object violates the default rule and also provides a reason to believe that the object-in-question is similar enough in this respect that it will also violate the rule, then infer that the object *does* violate the rule.

It is not similarity alone, but rather the availability of information explaining why the exception *is* an exception, and hinting that the object-in-question might fall under that explanation. And thirdly, our experiments concerning Benchmarks #1–4 asked subjects to differentiate *who* the agent solving the problem was supposed to be: a human (actually, the subject) or a robot. Interviews with pilot-study subjects indicated that this made a difference in the kinds of answers generated. (That is, people might take a different view about the inferences *they* would make with these default logic questions were *they* in the scenario described in the experiment from those inferences they would want or expect *an intelligent robot* to produce.) We had no *a priori* prediction or intuition about the human vs. robot dimension, but it seemed an interesting meta-cognitive issue to explore. The results were quite interesting and rather surprising, but cannot be gone deeply into here. We content ourselves with noting:

The Asimov Effect: People believe robots should be cautious (saying they "Can't tell") in cases whey they themselves would be willing to give a definite answer.

The Inheritance Reasoning Problems invoke a class-subclass hierarchy among different concepts and make assertions concerning typical properties of the concepts which occupy different positions in the hierarchy. These problems capture several essential questions concerning reasoning about classes, subclasses, and individuals: (a) how should properties – some of which are definitional and some of which are prototypical – be "inherited" by the next element down the hierarchy? and (b) how are conflicts to be resolved when, because of complex relationships, different values for the properties are available? As before, our interest here is to determine what people give as common sense answers to these sorts of problems. Looking at a graphical representation of Problems #11 and 12, one can discern that there is (what we call) a "hi node" vs. a "lo node", both of which are involved in inferences; and as well, one notices that the two

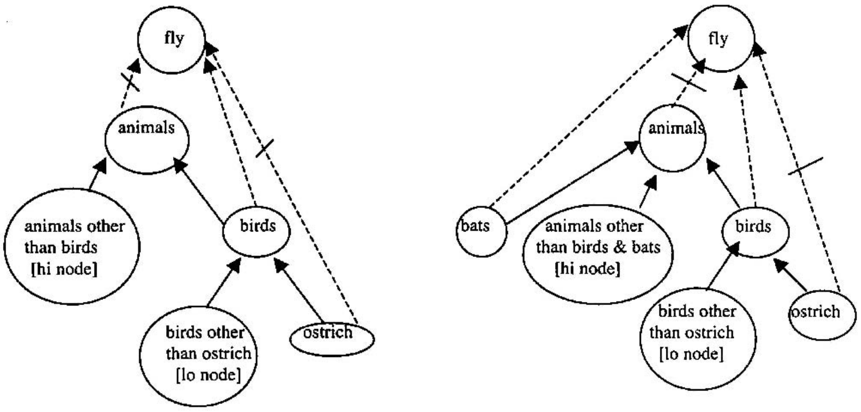

fly
animals
animals other than birds [hi node]
birds
birds other than ostrich [lo node]
ostrich
fly
bats
animals
animals other than birds & bats [hi node]
birds
birds other than ostrich [lo node]
ostrich

Figure 2. The structure of Benchmark Questions 11 and 12.

problems have the same structure except that #12 has an extra node (*bat*s) and two facts about this node.[8] Note that Benchmark Problem #12 does not request any inferences concerning this node; instead it merely defines a more complex network of relations within the hierarchy.

Approximately 90% of the subjects gave the AI answers for both nodes of Problem 11, and for the hi node of Problem 12. Our most striking finding was that subjects only gave the benchmark answer 53% of the time on Problem #12's low node. The other 47% are primarily incorrect answers rather than "can't tell" answers. That is, subjects would assert (using the birds-flying problem as an analogy) that birds-other-than-ostriches do not fly. This is incorrect, from the point of view of the AI answer, since it is expected that this node would inherit the default property from birds. After all, the only difference between Problem #11 and Problem #12 is the extra node on the high level in Problem #12, i.e., the addition that bats fly, and no inferences make use of this node. It seems peculiar that this addition would affect inferences about birds-other-than-ostriches. It reminds one of the classic default reasoning where the mere addition of a new premise (this time about bats) makes one retract a previously-drawn conclusion. One possibility is that explicitly mentioning that another subclass of animals exists and explicitly stating the status of some feature (*flying*) with respect to this class where this feature is used in the problem, subjects will thereby have two examples of subclasses that fly (*birds and bats*). They might note that these two subclasses that exist on the same level also have the same feature value, and they might apply this observation to the birds-other-than-ostriches node: namely, it should have the same feature value as the other class on its level.

A second study we ran on Benchmark Problems #11 and #12 investigated the effect of the kind of taxonomic categories being considered, namely *natural kind categories* vs. *artificial categories*. A second factor we manipulated concerned whether the problems included class-size information for the classes and subclasses that formed the inheritance hierarchy. For this initial study, we looked at just two cases. In the *no size* case, there was no class size information, and we presented the default rules in the form we have already been using. For the *class-size case*, the problem specified that the class size for each subclass was in the 20–80 range. The proportion of subject's answers that agreed with the AI answers was higher when subjects were reasoning about natural categories and lower when they were reasoning about artificial categories. And this was particularly notable when class size was mentioned.

The Artificial is Unnaturally Variable: People perceive artifacts and artificial classes as inherently more variable than natural kinds, and that this affects their willingness to ascribe default properties: the fact of merely mentioning class size apparently triggers this consideration of variability.

With respect to Problems #13–14 (the "Nixon problems"), we found that subjects gave the AI Answer to all questions except one: the question of whether Nixon is or is not a pacifist. The AI answer is that we cannot draw any conclusion, but almost half of the subjects say that in this sort of situation they *can* tell: about half of this half say he is a hawk and the others that he is a pacifist. (Recall that this is not a matter of world knowledge about Nixon or pacifists or Republicans, for our actual problems used different cover stories.) So even when faced with conflicting defaults subjects are willing to assert conclusions about Nixon's pacifist/hawk status. This may signal a case in which a different sort of default reasoning is taking place – more akin to the reasoning involved in Problems #1–#4 – namely, as a willingness to draw a conclusion even in an ambiguous situation (and perhaps tagging it as such). Perhaps in real-world situations, particularly in which decisions must be made or actions must be taken, a commonsense reasoner would draw *some* conclusion even in these 'conflict' nodes. One might even speculate that there aren't so many problems for which the world would tolerate a "can't tell" state in a problem-solver, and subjects' tendencies to go one way or the other on these conflict nodes may reflect a sensitivity to this reality. In the real world, after all, most people in fact say of the historical Nixon that he was "not really a Quaker".

6. CONCLUSIONS

There are two types of conclusions that we urge.

First type of conclusion: Unlike most other reasoning formalisms, nonmonotonic or default reasoning, including inheritance reasoning, *is* "psychologistic" – that is, it is defined only as what people do in circumstances where they are engaged in "commonsense reasoning". It follows from this that such fundamental issues as "what are the good nonmonotonic inferences?" or "what counts as 'relevance' to a default rule?", etc., are only discoverable by looking at *people and how they behave*. It is *not* a formal exercise to be discovered by looking at mathematical systems, nor is it to be decided by such formal considerations as "simplicity" or "computability", etc. Nor is it to be decided by researchers consulting only their own intuitions, or only those of their fellow researchers.

Second type of conclusion: The results of the experiments reported here point to some particular considerations that seem critical for nonmonontonic theories. We would urge theorists who are developing formal theories to incorporate our findings so as to make their theories more viable for the tasks that they are designed to solve.

ACKNOWLEDGMENTS

The authors would like to thank research assistants Siobjan Nearey and Matt Duncan, and the Department of Psychology at the University of Alberta for use of its research pool. They would also like to acknowledge the assistance of NSERC individual research grants, a joint NSERC equipment grant, and the University of Alberta Central Research Fund. Pelletier would like particularly to thank the participants of the Workshop on Paradigms of Cognition held at the University of Salzburg in August 2002, and the organizers of the Workshop, Hannes Leitgeb and Gerhard Schurz, for stimulating and helpful discussion of this material. We are very pleased that other researchers have decided to investigate these topics and have been quite surprised at the ingenuity with which these investigations have proceeded.

NOTES

[1] It is this feature, in fact, that makes many formal logicians question whether 'nonmonotonic logic' really is a logic. For, they say, an argument's being correct or valid means that whenever the premises are true, so is the conclusion. Thus if $\{P_1, P_2, \ldots P_n\} \therefore C$ is correct, then if $P_1, P_2, \ldots P_n$ are all true, so is C. But if this is the case, then how could $P_1, P_2, \ldots P_nP_{n+1}$ all be true and C not be true? After all, if $P_1, P_2, \ldots P_nP_{n+1}$ are all true, then in particular $P_1, P_2, \ldots P_n$ are true and hence C is true. So how *could* there be a nonmonotonic logic, these logicians ask. Isn't the concept of monotonicity built into our very notion of logic and correct reasoning?

[2] Pollock (1987, 1991) gives an account of defeasible reasoning which, however, does not take this viewpoint; instead it is in line with 'the knowledge representation viewpoint' we will outline below. He is also a good source for information on earlier accounts of defeasible reasoning.

[3] We said above that there is already a movement in the natural sciences to see their laws also as not having "strict and universal" laws. In this section we are merely pointing out that those who do not share this opinion of the natural sciences nonetheless hold it of the social sciences.

[4] For overviews see Bazerman (1986), Dawes (1988), Yates (1990); more detailed studies are discussed in Ross and Lepper (1980), Tomassini et al. (1982), Ashton and Ashton (1990).

[5] Although there are always people arguing against classical logic and for a "relevance logic" or a "fuzzy logic" or . . . , on the grounds that "this is the way people *actually* reason".

[6] Thanks to Joe Halpern for the example.

[7] For example, in the Hi Similarity version of the experiment, a problem specified where computer manuals are typically found: subjects were told that both the IBM and the Mac manuals were being reviewed by support staff because new software had been purchased. This assertion could be interpreted as *a reason why* the exception object (the Mac manual) was not where manuals typically were, *and also could be construed as giving a reason* to believe the object-in-question (the IBM manual) might also violate the rule.

[8] The solid lines indicate an exceptionless condition (e.g., all birds are animals), while the dashed lines indicate a default connection (e.g., birds normally fly). Lines through the dashed lines also indicate a default connection, but a negation-connection (ostriches normally do not fly).

REFERENCES

Anderson, A. and N. Belnap: 1975, *Entailment, I: The Logic of Relevance and Necessity*, Princeton University Press, Princeton.

Anderson, A., N. Belnap, M. Dunn, and R. Sylvan: 1992, *Entailment, II*, Princeton University Press, Princeton.

Ashton, R. and A. Ashton: 1990, 'Evidence-Responsiveness in Professional Judgment: Effects of Positive vs. Negative Evidence and Presentation Mode', *Organizational Behavior and Human Decision Processes* **46**, 1–19.

Bazerman, M.: 1986, *Judgment in Managerial Decision Making*, Wiley, New York.

Braine, M.: 1978, 'On the Relationship between the Natural Logic of Reasoning and Standard Logic', *Psychological Review* **85**, 1–21.

Carlson, G. and F. J. Pelletier (eds.): 1995, *The Generic Book*, University of Chicago Press, Chicago.

Cartwright, N.: 1983, *How the Laws of Nature Lie*, Oxford University Press, New York.

Cartwright, N.: 1989, *Nature's Capacities and Their Measurement*, Oxford University, New York.

Chater, N. and M. Oaksford: 1993, 'Logicism, Mental Models and Everyday Reasoning: Reply to Garnham', *Mind and Language* **8**, 72–89.

Clark, H. and S. Haviland: 1977, 'Comprehension and the Given-New Contract', in R. O. Freedle (ed.), *Discourse Production and Comprehension*, Ablex, Norwood, NJ, pp. 1–40.

Davis, E.: 1990, *Representations of Commonsense Knowledge*, Morgan Kaufmann, Palo Alto.

Davis, R.: 1984, 'Diagnostic Reasoning Based on Structure and Behavior', *Artificial Intelligence* **24**, 237–410.

Dawes, R.: 1988, *Rational Choice in an Uncertain World*, Hartcourt Brace, San Diego.

de Kleer, J. and B. Williams: 1987, 'Diagnosing Multiple Faults', *Artificial Intelligence* **32**, 97–130.

Elio, R. and F. J. Pelletier: 1993, 'Human Benchmarks on AI's Benchmark Problems', *Proceedings of the 15th Annual Cog. Sci. Society Conference*, Laurence Erlbaum, Hillsdale, pp. 406–411.

Elio, R. and F. J. Pelletier: 1996, 'On Reasoning with Default Rules and Exceptions', *Proceedings of the 18th Annual Cog. Sci. Society Conference*, Laurence Erlbaum, Hillsdale, pp. 131–136.

Evans, J.: 1987, *Bias in Human Reasoning: Causes and Consequences*, Laurence Erlbaum, Hillsdale.

Evans, R. and G. Gazdar: 1996, 'DATR: A Language for Lexical Knowledge Representation', *Computational Linguistics* **22**, 167–216.

Fetzer, J. and D. Nute: 1979, 'Syntax, Semantics, and Ontology: A Probabilistic Causal Calculus', *Synthese* **40**, 453–495.

Fodor, J. A.: 1975, *The Language of Thought*, Thomas Crowell, New York.

Fodor, J. A.: 1983, *The Modularity of Mind*, MIT Press, Cambridge.

Fodor, J. A. and Z. Pylyshyn: 1988, 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition* **28**, 3–71.

Frege, G.: 1884, *The Foundations of Arithmetic*, J. Austin, trans. (1950), Blackwell, Oxford.

Frege, G.: 1894, 'Review of Husserl *Philosophie der Arithmetic*', *Zeitschrift für Philosophie und philosophische Kritik* **103**, 313–332. Reprinted in P. Geach and M. Black: 1952, *Translations from the Philosophical Writings of Gottlob Frege*, Blackwells, Oxford, pp. 79–85.

Friedman, M.: 1957, *The Theory of the Consumption Function*, Princeton University Press, Princeton.

Garnham, A.: 1993, 'Is Logicist Cognitive Science Possible?', *Mind and Language* **8**, 49–71.

Ginsberg, M.: 1987, 'Introduction', in M. Ginsberg (ed.), *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann, Los Altos, CA.

Goebel, R. and S. Goodwin: 1987, 'Applying Theory Formation to the Planning Problem', in F. Brown (ed.), *The Frame Problem in Artificial Intelligence*, University of Kansas Press, Lawrence, pp. 207–232.

Good, I. J.: 1961/1962, ‘A Causal Calculus (I and II)’, *British Journal for Philosophy of Science* **11**, 305–318, **12**, 43–51.

Grice, H. P.: 1975, ‘Logic and Conversation’, in P. Cole and J. Morgan (eds.), *Syntax and Semantics, Vol. 3: Speech Acts*, Academic Press, New York, pp. 41–58.

Grice, H. P.: 1978, ‘Further Notes on Logic and Conversation’, in P. Cole (ed.), *Syntax and Semantics, Vol. 9: Pragmatics*, Academic Press, New York, pp. 113–128.

Grice, H. P.: 1981, ‘Presupposition and Conversational Implicature’, in P. Cole (ed.), *Radical Pragmatics*, Academic Press, New York, pp. 183–198.

Hammond, J.: 1988, ‘How Different are Friedman and Hicks on Method?’, *Oxford Economic Papers* **40**, 392–394.

Hanks, S. and D. McDermott: 1986, ‘Default Reasoning, Nonmonotonic Logics, and the Frame Problem’, *Proceedings of AAAI-86*, pp. 328–333.

Hanks, S. and D. McDermott: 1987, ‘Nonmonotonic Logic and Temporal Projection’, *Artificial Intelligence* **33**, 379–412.

Hausman, D.: 1989, ‘Economic Methodology in a Nutshell’, *Journal of Economic Perspectives* **3**, 115–127.

Henle, M.: 1962, ‘On the Relation between Logic and Thinking’, *Psychological Review* **69**, 366–378.

Hirsch, A. and N. de Marchi: 1986, ‘Making a Case When a Theory is Unfalsifiable: Friedman’s Monetary History’, *Economics and Philosophy* **2**, 1–21.

Jachowicz, P. and R. Goebel: 1997, ‘Describing Plan Recognition as Non-Monotonic Reasoning and Belief Revision’, *Proceedings of the Australian Joint Conference on AI*, pp. 236–245.

Janssen, M. and Y. Tan: 1991, ‘Why Friedman’s Non-Monotonic Reasoning Defies Hempel’s Covering Law Model’, *Synthese* **86**, 255–284.

Janssen, M. and Y. Tan: 1992, ‘Friedman’s Permanent Income Hypothesis as an Example of Diagnostic Reasoning’, *Economics and Philosophy* **8**, 23–49.

Johnson-Laird, P. N.: 1983, *Mental Models*, Harvard University Press, Cambridge.

Krifka, M., F. J. Pelletier, G. Carlson, A. ter Meulen, G. Chierchia, and G. Link: 1995, ‘Genericity: An Introduction’, in G. Carlson and F. J. Pelletier (eds.), pp. 1–124.

Lakoff, G.: 1972, ‘Linguistics and Natural Logic’, in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Reidel, Dordrecht, pp. 232-296.

Lakoff, G.: 1973, ‘Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts’, *Journal of Philosophical Logic* **2**, 458–508.

Lascarides, A. and N. Asher: 1993, ‘Temporal Interpretation, Discourse Relations and Common Sense Entailment’, *Linguistics and Philosophy* **16**, 437–493.

Lascarides, A., T. Briscoe, N. Asher, and A. Copestake: 1996, ‘Order Independent and Persistent Typed Default Unification’, *Linguistics and Philosophy* **19**, 1–89.

Levesque, H. and G. Lakemeyer: 2001, *The Logic of Knowledge Bases*, MIT Press, Cambridge.

Levinson, S.: 1983, *Pragmatics*, Cambridge University Press, Cambridge.

Lewis, D.: 1973, *Counterfactuals*, Cambridge University Press, Cambridge.

Lifschitz, V.: 1989, ‘25 Benchmark Problems in Nonmonotonic Reasoning’, in M. Reinfrank, J. de Kleer, and M. Ginsberg (eds.), *Nonmonotonic Reasoning*, Vol. 2.0, Springer Verlag, Berlin, pp. 202–219.

Macnamara, R.: 1986, *A Border Dispute: The Place of Logic in Psychology*, MIT Press, Cambridge.

McCarthy, J.: 1986, ‘Applications of Circumscription to Formalizing Common Sense Knowledge’, *Artificial Intelligence* **28**, 89–116.

McCarthy, J. and P. Hayes: 1969, 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', in B. Meltzer and D. Michie (eds.), *Machine Intelligence*, Vol. 4, University of Edinburgh Press, Edinburgh, pp. 463–502.

McDermott, D.: 1982, 'A Temporal Logic for Reasoning about Processes and Plans', *Cognitive Science* **6**, 101–155.

Medin, D. and E. Smith: 1984, 'Concepts and Concept Formation', *Annual Review of Psychology* **35**, 113–138.

Moore, R.: 1985, 'Semantical Considerations on Nonmonotonic Logic', *Artificial Intelligence* **25**, 75–94.

Nisbett, R., d. Krantz, D. Jepson, and Z. Kunda: 1983, 'The Use of Statistical Heuristics in Everyday Inductive Reasoning', *Psychological Review* **90**, 339–363.

Oaksford, M. and N. Chater: 1991, 'Against Logicist Cognitive Science', *Mind and Language* **6**, 1–38.

Pearl, J.: 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufman, San Mateo.

Pearl, J.: 2000, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge.

Pelletier, F. J. and R. Elio: 1997, 'What Should Default Reasoning be, by Default?', *Computational Intelligence* **13**, 165–187.

Pelletier, F. J. and R. Elio: 2003, 'Logic and Cognition: Human Performance in Default Reasoning', in P. Gärdenfors, J. Wolenski, K. Kijania-Placet (eds.), *In the scope of Logic, Methodology, and Philosophy of Science, Vol. I*, Kluwer Academic Publishers, Dordrecht, pp. 137–154.

Pollock, J.: 1987, 'Defeasible Reasoning, *Cognitive Science* **11**, 481–518.

Pollock, J.: 1991, 'A Theory of Defeasible Reasoning', *International Journal of Intelligent Systems* **6**, 33–54.

Poole, D., R. Aleliunas, and R. Goebel: 1987, 'Theorist: A Logical Reasoning System for Defaults and Diagnosis', in N. Cercone and G. McCalla (eds.), *The Knowledge Frontier: Essays in the Representation of Knowledge*, Springer Verlag, New York, pp. 331–352.

Putnam, H.: 1970, 'Is Semantics Possible?, in H. Kiefer and M. Munitz (eds.), *Language, Belief, and Metaphysics*, SUNY Press, Binghamton pp. 50–63.

Putnam, H.: 1975, 'The Meaning of 'Meaning'', in K. Gunderson (ed.), *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science, University of Minnesota Press, Minneapolis, pp. 131–193.

Pylyshyn, Z.: 1984, *Computation and Cognition: Towards a Foundation for Cognitive Science*, Bradford Books, Montgomery, VT.

Reiter, R.: 1978, 'On Reasoning by Default', *Proceedings of TINLAP-2*, Association for Computational Linguistics, University of Illinois, pp. 210–218.

Reiter, R.: 1980, 'A Logic for Default Reasoning', *Artificial Intelligence* **13**, 81–132.

Reiter, R.: 1987, 'A Theory of Diagnosis from First Principles', *Artificial Intelligence* **32**, 97–130.

Reiter, R.: 1992, 'What Should a Database Know?', *Journal of Logic Programming* **14**, 127–153.

Rosch, E.: 1978, 'Principles of Categorization', in E. Rosch and B. B. Lloyd (eds.), *Cognition and Categorization*, Lawrence Erlbaum, Hillsdale, pp. 27–48.

Ross, L. and M. Lepper: 1980, 'The Perseverance of Beliefs: Empirical and Normative Considerations', in R. Shweder (ed.), *New Directions for Methodology of Social and Behavioral Science: Fallible Judgment in Behavioral Research*, Jossey-Bass, San Francisco, pp. 17–36.

Ross, W. D.: 1930, *A Theory of the Right and the Good*, Oxford University Press, Oxford.

Salmon, W.: 1990, *Four Decades of Scientific Explanation*, University of Minnesota Press, Minneapolis.

Schubert, L. and F. J. Pelletier: 1987, ‘Problems in the Representation of the Logical Form of Generics, Bare Plurals, and Mass Terms’, in E. LePore (ed.), *New Directions in Semantics*, Academic Press, New York, pp. 385–451.

Searle, J.: 1964, ‘How to Derive ‘Ought’ from ‘Is”, *Philosophical Review* **73**, 43–58.

Selman, B. and H. Kautz: 1989, ‘The Complexity of Model-Preference Default Theories’, in M. Reinfrank, J. de Kleer, and M. Ginsberg (eds.), *Nonmonotonic Reasoning*, Springer Verlag, Berlin, pp. 115–130.

Shoham, Y.: 1990, ‘Nonmonotonic Temporal Reasoning and Causation’, *Cognitive Science* **14**, 213–252.

Smith, E. and D. Osherson: 1984, ‘Conceptual Combination with Prototype Concepts’, *Cognitive Science* **8**, 337–361.

Sperber, D. and D. Wilson, 1986, *Relevance: Communication and Cognition*, Blackwell, Oxford.

Suppes, P.: 1970, *A Probabilistic Theory of Causation*, North-Holland, Amsterdam.

Thomason, R.: 1997, ‘Nonmonotonicity in Linguistics’, in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam, pp. 777–831.

Tobin, J.: 1971, ‘Money and Income: Post Hoc Ergo Propter Hoc’, in J. Tobin (ed.), *Essays in Macroeconomics*, University of Chicago Press, Chicago, pp. 497–514.

Tomassini, L., I. Solomon, M. Romney, and J. Krogstad: 1982, ‘Calibration of Auditors’ Probabilistic Judgments: Some Empirical Evidence’, *Organizational Behavior and Human Performance* **30**, 137–151.

Tversky, A. and D. Kahneman: 1983, ‘Extensional *versus* Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment’, *Psychological Review* **90**, 293–315.

Yates, J. F.: 1990, *Judgment and Decision Making*, Prentice Hall, Englewood Cliffs.

Francis Jeffry Pelletier
Department of Philosophy
University of Alberta
Canada T6G 2E5
E-mail: jeffp@cs.ualberta.ca

Renée Elio
Department of Computing Science
University of Alberta
Edmonton, Alberta
Canada T6G 2H1
E-mail: ree@cs.ualberta.ca